Hepatoprotective effects of systemic ER activation

Spheroid RNA-seq - Differential expression analysis

Christian Sommerauer & Carlos Gallardo

25 July, 2023

Load packages

```
source("code/00_helper_functions.R")
library(tidyverse)
library(DESeq2)
library(biomaRt)
library(ggvenn)
```

Load featureCount table

```
count_table <- read.table("data/raw_fq_featurecounts_inhibitors_spheroids.txt", header=T) %>% drop_na()

names(count_table) <- gsub("hisat2_aligned\\.|\\.bam", "", names(count_table))

count_table.2 <- count_table  %>% tibble::column_to_rownames("Geneid") %>% dplyr::select(2:9)
head(count_table.2)
```

```
##                 FFA.control.1_S10 FFA.control.2_S11 FFA.control.3_S12
## ENSG00000160072               160               142                48
## ENSG00000279928                 0                 0                 0
## ENSG00000228037                 0                 0                 0
## ENSG00000142611                 3                 0                 0
## ENSG00000284616                 0                 0                 0
## ENSG00000157911               200               211                31
##                 TEADap.inh.1_S13 TEADap.inh.2_S14 TEADsf.inh.1_S16
## ENSG00000160072              152              133              120
## ENSG00000279928                0                0                0
## ENSG00000228037                0                0                0
## ENSG00000142611                3                4                1
## ENSG00000284616                0                0                0
## ENSG00000157911              218              198              208
##                 TEADsf.inh.2_S17 TEADsf.inh.3_S18
## ENSG00000160072              125              247
## ENSG00000279928                0                0
## ENSG00000228037                0                1
## ENSG00000142611                6                1
## ENSG00000284616                0                0
## ENSG00000157911              240              388
```

Design metatables

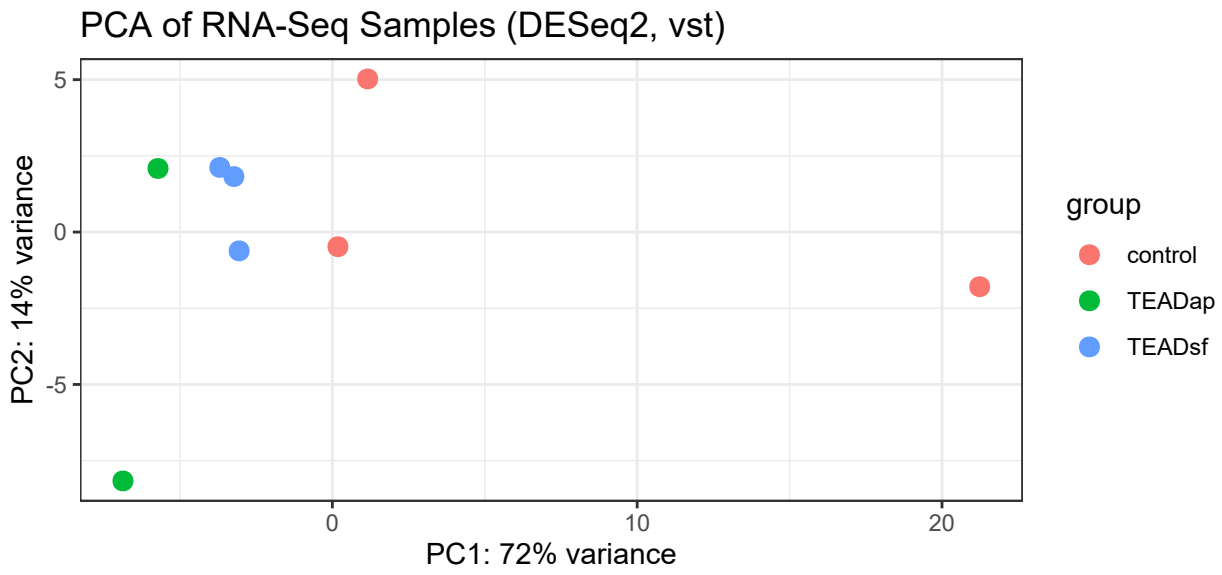
```
sample_name_inhibitors <- names(count_table.2)
replicate_inhibitors <- c(1,2,3,1,2,1,2,3)
Condition_inhibitors <- c(rep("control",3), rep("TEADap",2), rep("TEADsf",3))
meta_table_inhibitors <- data.frame(Sample=sample_name_inhibitors, Condition=Condition_inhibitors, Replicate=replicate_inhibitors)
write.table(meta_table_inhibitors, "results/spheroid_meta_table.txt", sep="\t", quote=F)
```

Export TPM-normalized count table

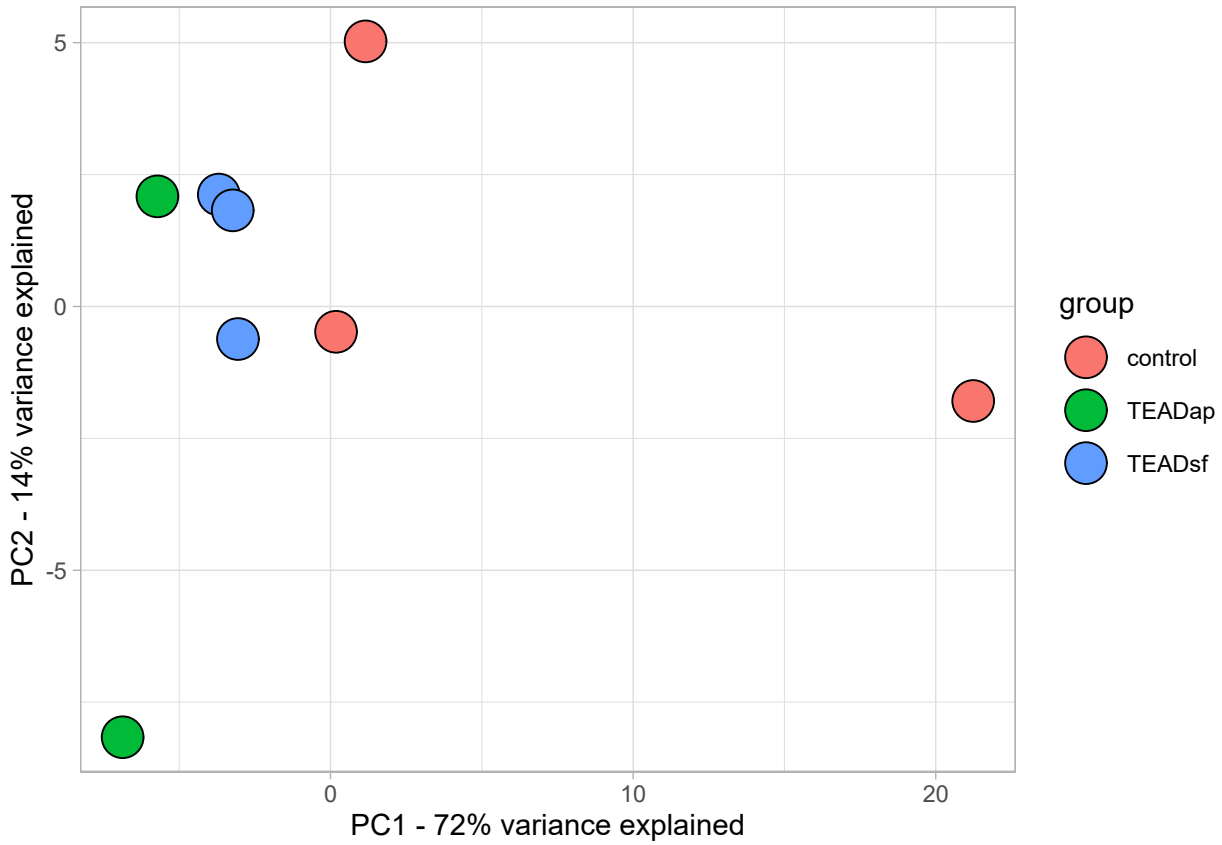
```
# normalize to TPM
count_table_TPM <- normalizeData(x=count_table.2, method = "TPM", len = count_table$Length)

count_table_TPM_mean <- groupTransform(x=count_table_TPM,
               group.lbls = meta_table_inhibitors$Condition,
               FUN=function(x) apply(x, 1, mean)) %>%
               tibble::rownames_to_column("ensembl_gene_id")

# move the rownames to column
count_table_TPM <- count_table_TPM %>%
             tibble::rownames_to_column("ensembl_gene_id")

# Add the external_gene_names
#listEnsemblArchives()
mart <- useMart(biomart = "ensembl", dataset = "hsapiens_gene_ensembl", host="https://oct2022.archive.ensembl.org") #, host = "https://oct2022.arc
annoData <- getBM(attributes=c("ensembl_gene_id","external_gene_name", "description"),mart=mart)

export_TPM_normTables <- list(count_table_TPM=count_table_TPM,
                              count_table_TPM_mean=count_table_TPM_mean)

export_TPM_normTables_single <- inner_join(export_TPM_normTables$count_table_TPM, y = annoData, by = 'ensembl_gene_id')
export_TPM_normTables_mean <- inner_join(export_TPM_normTables$count_table_TPM_mean, y = annoData, by = 'ensembl_gene_id')

write.table(export_TPM_normTables_single, "results/spheroid_TPM_norm_counts.txt", sep="\t", quote=F, row.names = F)
write.table(export_TPM_normTables_mean, "results/spheroid_TPM_norm_counts_mean.txt", sep="\t", quote=F, row.names = F)
```

Run DESeq2 function and plot a PCA

```
ds_data_inhibitors <- DESeqDataSetFromMatrix(countData = count_table.2,
                                colData = meta_table_inhibitors,
                                design = ~ 0 + Condition)
ds_data_inhibitors <- estimateSizeFactors(ds_data_inhibitors)
ds_data_inhibitors <- DESeq(ds_data_inhibitors)

# filtering to minimum 15 counts in at least 8 samples
ds_data_inhibitors.filt <- ds_data_inhibitors[rowSums(counts(ds_data_inhibitors, normalized=TRUE) >= 10 ) >= 2, ]

vsd_inhibitors <- vst(ds_data_inhibitors.filt)

PCA.to.plot_inhibitors <- plotPCA(vsd_inhibitors, intgroup=c("Condition")) + theme_bw() +  ggtitle("PCA of RNA-Seq Samples (DESeq2, vst)")
PCA.to.plot_inhibitors
```




```
ggplot(PCA.to.plot_inhibitors$data, aes(x=PC1, y=PC2, fill=group), color="black") +
  geom_point(shape=21, size=7) +
  theme_light() +
  ylab("PC2 - 14% variance explained") + # took the number from the original plotPCA function
  xlab("PC1 - 72% variance explained")
```




```
DESeq2_DEGs_inhibitors <- list()
# comparisons
DESeq2_DEGs_inhibitors$unfilt <- list(
  TEADap_vs_control = results(ds_data_inhibitors.filt, contrast = c('Condition', 'TEADap', 'control')),
  TEADsf_vs_control = results(ds_data_inhibitors.filt, contrast = c('Condition', 'TEADsf', 'control')))

names(DESeq2_DEGs_inhibitors$unfilt)
```

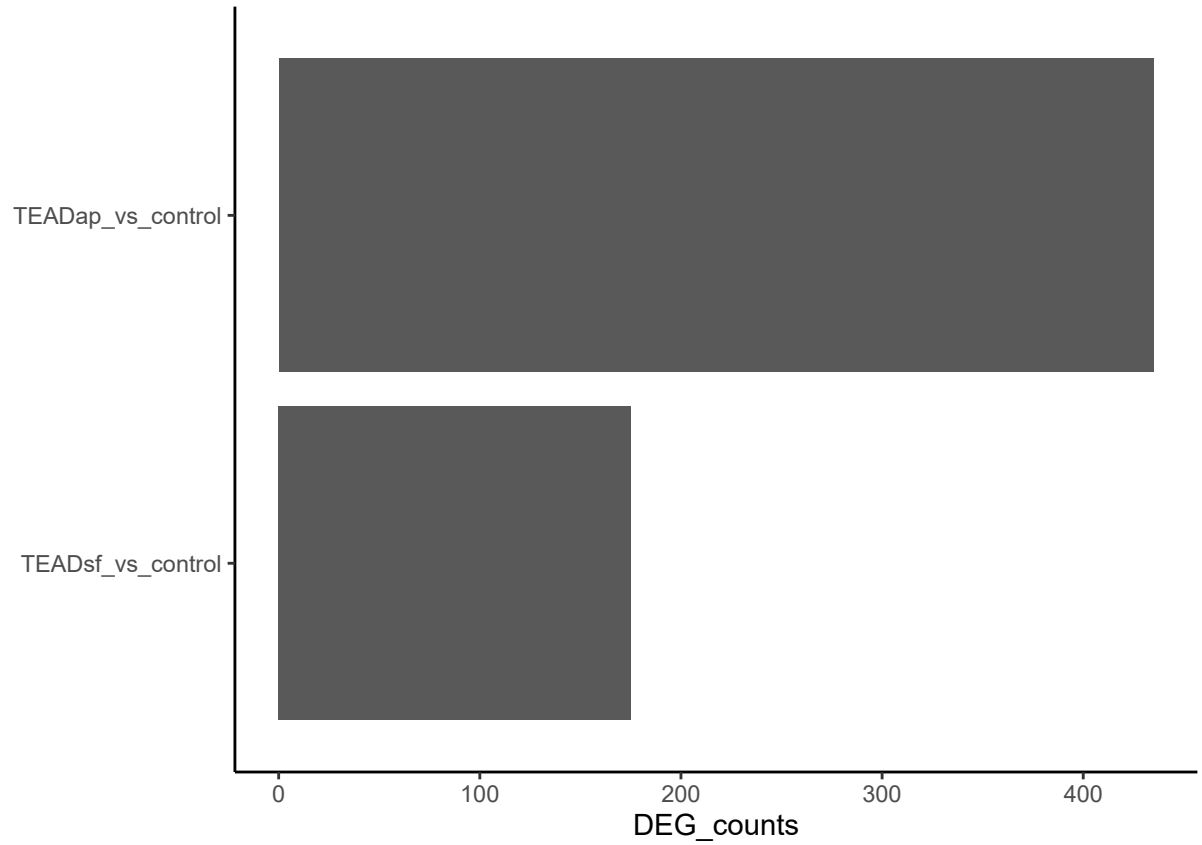
```
## [1] "TEADap_vs_control" "TEADsf_vs_control"
```

```
# add annotation to DEG lists
DESeq2_DEGs_inhibitors$unfilt <- lapply(DESeq2_DEGs_inhibitors$unfilt, as.data.frame)
DESeq2_DEGs_inhibitors$unfilt <- lapply(DESeq2_DEGs_inhibitors$unfilt, rownames_to_column, var = 'ensembl_gene_id')

listEnsemblArchives()
```

```
##              name     date                                url version
## 1  Ensembl GRCh37 Feb 2014          https://grch37.ensembl.org  GRCh37
## 2     Ensembl 110 Jul 2023 https://jul2023.archive.ensembl.org     110
## 3     Ensembl 109 Feb 2023 https://feb2023.archive.ensembl.org     109
## 4     Ensembl 108 Oct 2022 https://oct2022.archive.ensembl.org     108
## 5     Ensembl 107 Jul 2022 https://jul2022.archive.ensembl.org     107
## 6     Ensembl 106 Apr 2022 https://apr2022.archive.ensembl.org     106
## 7     Ensembl 105 Dec 2021 https://dec2021.archive.ensembl.org     105
## 8     Ensembl 104 May 2021 https://may2021.archive.ensembl.org     104
## 9     Ensembl 103 Feb 2021 https://feb2021.archive.ensembl.org     103
## 10    Ensembl 102 Nov 2020 https://nov2020.archive.ensembl.org     102
## 11    Ensembl 101 Aug 2020 https://aug2020.archive.ensembl.org     101
## 12    Ensembl 100 Apr 2020 https://apr2020.archive.ensembl.org     100
## 13     Ensembl 99 Jan 2020 https://jan2020.archive.ensembl.org      99
## 14     Ensembl 98 Sep 2019 https://sep2019.archive.ensembl.org      98
## 15     Ensembl 97 Jul 2019 https://jul2019.archive.ensembl.org      97
## 16     Ensembl 96 Apr 2019 https://apr2019.archive.ensembl.org      96
## 17     Ensembl 95 Jan 2019 https://jan2019.archive.ensembl.org      95
## 18     Ensembl 94 Oct 2018 https://oct2018.archive.ensembl.org      94
## 19     Ensembl 93 Jul 2018 https://jul2018.archive.ensembl.org      93
## 20     Ensembl 80 May 2015 https://may2015.archive.ensembl.org      80
## 21     Ensembl 77 Oct 2014 https://oct2014.archive.ensembl.org      77
## 22     Ensembl 75 Feb 2014 https://feb2014.archive.ensembl.org      75
## 23     Ensembl 54 May 2009 https://may2009.archive.ensembl.org      54
```

```
##    current_release
## 1
## 2                *
## 3
## 4
## 5
## 6
## 7
## 8
## 9
## 10
## 11
## 12
## 13
## 14
## 15
## 16
## 17
## 18
## 19
## 20
## 21
## 22
## 23
```

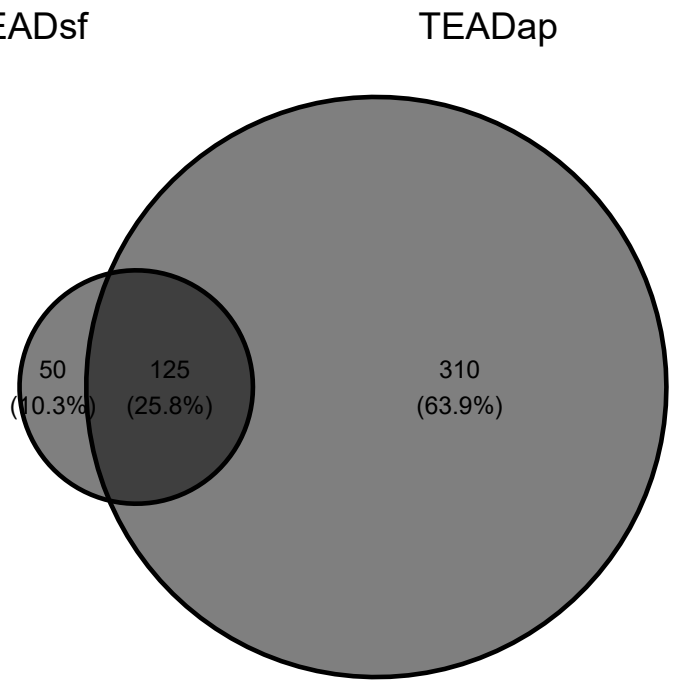
```
mart <- useMart(biomart = "ensembl", dataset = "hsapiens_gene_ensembl", host = "https://oct2022.archive.ensembl.org")
# use the getAnnotation function to obtain relevant features for ensembl GRCh38.p13.
annoData <- getBM(attributes=c("ensembl_gene_id","external_gene_name", "chromosome_name", "gene_biotype", "description"),mart=mart)

DESeq2_DEGs_inhibitors$unfilt <- lapply(DESeq2_DEGs_inhibitors$unfilt, inner_join, y = annoData, by = 'ensembl_gene_id')

DEG_counts_inhibitors <- vector()
for (i in 1:2) {
DESeq2_DEGs_inhibitors$filt[[i]] <- DESeq2_DEGs_inhibitors$unfilt[[i]] %>% filter(abs(log2FoldChange) > 0.585 & padj<0.05)
DEG_counts_inhibitors[i] <- nrow(DESeq2_DEGs_inhibitors$filt[[i]])
}

names(DESeq2_DEGs_inhibitors$filt) = names(DESeq2_DEGs_inhibitors$unfilt)

DEG_counts_inhibitors <- data.frame(Condition = names(DESeq2_DEGs_inhibitors$filt), DEG_counts = DEG_counts_inhibitors)

ggplot(DEG_counts_inhibitors, aes(x=DEG_counts, y=factor(Condition, levels=rev(names(DESeq2_DEGs_inhibitors$unfilt))))) +
  geom_col() +
  theme_classic() +
  ylab("")
```




```
saveRDS(DESeq2_DEGs_inhibitors, "results/Spheroid_inhibitors_DEGlists.rds")

write.table(DESeq2_DEGs_inhibitors$unfilt$TEADap_vs_control, "results/TEADap_vs_control_unfilt_DESeq2.txt", sep="\t", row.names=F, quote=F)
write.table(DESeq2_DEGs_inhibitors$unfilt$TEADsf_vs_control, "results/TEADsf_vs_control_unfilt_DESeq2.txt", sep="\t", row.names=F, quote=F)
write.table(DESeq2_DEGs_inhibitors$filt$TEADap_vs_control, "results/TEADap_vs_control_DEGs_DESeq2.txt", sep="\t", row.names=F, quote=F)
write.table(DESeq2_DEGs_inhibitors$filt$TEADsf_vs_control, "results/TEADsf_vs_control_DEGs_DESeq2.txt", sep="\t", row.names=F, quote=F)
```

Venn diagram TEADsf & TEADap DEGs

```
DEGs_TEADap <- DESeq2_DEGs_inhibitors$filt$TEADap_vs_control %>%
  pull(ensembl_gene_id) %>%
  unique()

DEGs_TEADsf <- DESeq2_DEGs_inhibitors$filt$TEADsf_vs_control %>%
  pull(ensembl_gene_id) %>%
  unique()

venn <- list(TEADap = DEGs_TEADap,
             TEADsf = DEGs_TEADsf)

# create venn diagram
ggvenn(venn, c("TEADsf", "TEADap"),
       auto_scale=TRUE,
       fill_color = c("black", "black"))
```



SessionInfo

```
sessionInfo()
```

```
## R version 4.2.1 (2022-06-23 ucrt)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 19045)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=English_United States.utf8
## [2] LC_CTYPE=English_United States.utf8
## [3] LC_MONETARY=English_United States.utf8
## [4] LC_NUMERIC=C
## [5] LC_TIME=English_United States.utf8
##
## attached base packages:
## [1] grid      stats4    stats     graphics  grDevices utils     datasets
## [8] methods   base
##
## other attached packages:
##  [1] ggvenn_0.1.10               biomaRt_2.54.1
##  [3] DESeq2_1.38.1               SummarizedExperiment_1.28.0
##  [5] Biobase_2.58.0              MatrixGenerics_1.10.0
##  [7] matrixStats_0.63.0          GenomicRanges_1.50.1
##  [9] GenomeInfoDb_1.34.9         IRanges_2.32.0
## [11] S4Vectors_0.36.0            BiocGenerics_0.44.0
## [13] lubridate_1.9.2             forcats_1.0.0
## [15] stringr_1.5.0               dplyr_1.1.0
## [17] purrr_1.0.1                 readr_2.1.4
## [19] tidyr_1.3.0                 tibble_3.2.1
## [21] ggplot2_3.4.2               tidyverse_2.0.0
##
## loaded via a namespace (and not attached):
##  [1] bitops_1.0-7           bit64_4.0.5            filelock_1.0.2
##  [4] RColorBrewer_1.1-3     progress_1.2.2         httr_1.4.6
##  [7] tools_4.2.1            utf8_1.2.2             R6_2.5.1
## [10] DBI_1.1.3              colorspace_2.0-3       withr_2.5.0
## [13] tidyselect_1.2.0       prettyunits_1.1.1      bit_4.0.5
## [16] curl_5.0.1             compiler_4.2.1         cli_3.4.1
## [19] xml2_1.3.5             DelayedArray_0.24.0    labeling_0.4.2
## [22] scales_1.2.1           rappdirs_0.3.3         digest_0.6.30
## [25] rmarkdown_2.23         XVector_0.38.0         pkgconfig_2.0.3
```

```
## [28] htmltools_0.5.5        highr_0.10             dbplyr_2.3.3
## [31] fastmap_1.1.0          rlang_1.1.1            rstudioapi_0.15.0
## [34] RSQLite_2.2.19         generics_0.1.3         farver_2.1.1
## [37] BiocParallel_1.32.3    RCurl_1.98-1.9         magrittr_2.0.3
## [40] GenomeInfoDbData_1.2.9 Matrix_1.5-3           Rcpp_1.0.9
## [43] munsell_0.5.0          fansi_1.0.3            lifecycle_1.0.3
## [46] stringi_1.7.8          yaml_2.3.7             zlibbioc_1.44.0
## [49] BiocFileCache_2.6.1    blob_1.2.4             parallel_4.2.1
## [52] crayon_1.5.2           lattice_0.20-41        Biostrings_2.66.0
## [55] annotate_1.76.0        hms_1.1.3              KEGGREST_1.38.0
## [58] locfit_1.5-9.6         knitr_1.43             pillar_1.9.0
## [61] geneplotter_1.76.0     codetools_0.2-19       XML_3.99-0.12
## [64] glue_1.6.2             evaluate_0.21          png_0.1-8
## [67] vctrs_0.6.2            tzdb_0.4.0             gtable_0.3.3
## [70] cachem_1.0.6           xfun_0.39              xtable_1.8-4
## [73] AnnotationDbi_1.60.2   memoise_2.0.1          timechange_0.2.0
```